Manay Patel - S2AND on Patent Data README
Fall 2021

Setting Up Directories:

Move all the files and directories into the main folder for patent data (say patentsview/).
patentsview/ must contain the patent_paper.json file and folders corresponding to training and testing files
For example, patentsview/ could look like this:

patentsview/:
- -patent_paper.json
- -common_charac/: [FOR TRAINING]
  - -patent_signatures.json
  - -patent_clusters.json
  - -patent_specter.pickle
- -als_common/: [FOR TESTING]
  - -patent_signatures.json
  - -patent_clusters.json

Brief Summary of Running S2AND Model:

The model consists primarily of three class objects:

- ANDData class: This is the main class that reads and processes all of the datasets.

- PairwiseModeler class: This class builds upon ANDData object and creates an object which does pairwise functions.

- Clusterer class: This class builds upon ANDData object and creates an object which does clustering functions.

To save any of these objects, AllenAi shows a way to save the object in a pickle file which can be opened to perform any of the object's functions.

Modifications made to S2AND Model:

There are two additional modes added to S2AND: "only_train_val_split" and "only_test" which are specified while creating the ANDData object for train and test datasets respectively.

These modes ensure that the data set is not unnecessarily split into more than necessary splits.

Example run files:

To run all training and testing under one dataset, check the train_test_together.py script.
To run training and testing under different datasets, check the run_train_test_separate.py script.
To see how to make predictions using S2AND, check the testPredict.py script.